Journal of Attention Disorders Vol. 12(1):4-14 (2008)
ISSN: 1087-0547
doi:10.1177/1087054708319345

Evidence, Interpretation, and Qualification From Multiple Reports of Long-Term Outcomes in the Multimodal Treatment Study of Children With ADHD (MTA), Part I: Executive Summary

James Swanson, L. Eugene Arnold, Helena Kraemer, Lily Hechtman, Brooke Molina, Stephen Hinshaw, Benedetto Vitiello, Peter Jensen, Ken Steinhoff, Marc Lerner, Laurence Greenhill, Howard Abikoff, Karen Wells, Jeffery Epstein, Glen Elliott, Jeffrey Newcorn, Betsy Hoza, and Timothy Wigal *(MTA Cooperative Group)*

Objective:

To review the primary and secondary findings from the Multimodal Treatment study of ADHD (MTA) published over the past decade as three sets of articles.

Method:

In a two-part article—Part I: Executive Summary (without distracting details) and Part II: Supporting Details (with additional background and detail required by the complexity of the MTA)—we address confusion and controversy about the findings.

Results:

We discuss the basic features of the gold standard used to produce scientific evidence, the randomized clinical trial, for which was used to contrast four treatment conditions: medication management alone (MedMgt), behavior therapy alone (Beh), the combination of these two (Comb), and a community comparison of treatment "as usual" (CC). For each of the three assessment points we review three areas that we believe are important for appreciation of the findings: definition of evidence from the MTA, interpretation of the serial presentations of findings at each assessment point with a different definition of long-term, and qualification of the interim conclusions about long-term effects of treatments for ADHD.

Conclusion:

We discuss the possible clinical relevance of the MTA and present some practical suggestions based on current knowledge and uncertainties facing families, clinicians, and investigators regarding the long-term use of stimulant medication and behavioral therapy in the treatment of children with ADHD. *(J. of Att. Dis. 2008; 12(1) 4-14)*

A. Introduction and Purpose of Part I (Executive Summary)

1. Introduction

The main findings from the Multimodal Treatment study of ADHD (MTA) are described in eight articles published over the past decade. The first two articles were published in the *Archives of General Psychiatry* about the 14-month treatment phase (MTA Cooperative Group, 1999a, 1999b); the next two articles in *Pediatrics* about the initial 10-month follow-up phase (MTA Cooperative Group, 2004a, 2004b); and the most recent four articles in the *Journal of the American Academy of Child and Adolescent Psychiatry* about the subsequent 12-month follow-

up phase (Jensen et al., 2007; Swanson, Elliott, et al., 2007; Swanson, Hinshaw, et al., 2007; Molina et al., 2007). The serial reports of complex findings in different journals may make it difficult for readers to integrate the main findings, partly because a summary of the information is not accessible in one place. Therefore, one of our main objectives is to review and integrate the findings from these three sets of articles and make this information accessible to those interested in the MTA.

We will use the terms "primary" and "secondary" to distinguish the articles within each set. The first, or *primary*, article in each set presents the findings (up to a point in time) based on the logic and methodology of our randomized clinical trial (RCT) design and narrowly focused and rigorous intent-to-treat (ITT) method of analysis. A defining feature of the RCT approach is random assignment of participants to treatment conditions, and a defining feature of the ITT approach is the evaluation of all participants randomized and assigned to the well-defined treatments whether or not they were accepted or implemented as planned. We consider the findings from the primary articles based on RCT principles and ITT analyses to represent the scientific evidence from the MTA. The next, or *secondary*, article in each set presents interpretations of findings based on a variety of sophisticated and often complicated analyses that supplement the primary analyses. These include the systematic explorations of mediators and moderators of treatment effects and the evaluation of transition of the RCT into an observational follow-up study, which introduces additional conceptual and statistical issues for consideration. These should be distinguished from the many *additional* articles that we have published (see the appendix to Part II) using exploratory analyses to suggest possible qualification to the primary and secondary findings. We believe these articles have advanced our knowledge about treatment of ADHD, but we realize they may have also contributed to confusion and controversy, since the varied purposes of these additional articles were to suggest many possible qualifications about the findings reported in the primary and secondary articles of the MTA.

We believe that discussion of the findings from the MTA should be based on a firm understanding and appreciation of the strengths and limitations of the design and methods that are presented and described in the primary and secondary articles. We hope to dispel some confusion and resolve some controversies by a thorough but focused review of these eight main articles. This should provide the factual basis for evidence-based interpretations of some of the current uncertainties facing families, clinicians, and investigators about the long term use of stimulant medication and behavioral therapy for treatment of children with ADHD.

2. Purpose

After each main set of publications, questions should be expected about the interim findings of a large, publicly funded, national study that is still in progress. Even among MTA investigators, there has been open discussion of differences in opinion about emphasis, importance, and interpretation of the initial and follow-up findings, and commentaries with a variety of viewpoints have appeared in scientific journals as well as in popular media. Because discussion of this is a valuable part of the dissemination of findings, our primary purpose is to identify some relevant issues about the evidence, interpretation, and qualification of the findings that will help address some of the confusion and controversies about the MTA findings through the 36-month assessment.

Integrating the main findings of this very complex series of reports is not a simple task. Our first attempt was too long and detailed, so we separated it into two parts. Here in the first part (the Executive Summary), we highlight a few key aspects of the three sets of articles, and in the second part (Supporting Details) we will provide further explanations and discussions that are

important but heavy with excruciating detail. In parallel sections of these two articles, for each of the three sets of publications we will (a) define what we consider to be evidence from the MTA, (b) address confusion and controversies that have been generated by some interpretations of the primary and secondary findings the MTA, and (c) discuss what we consider to be important qualifications suggested by additional articles intended to supplement the evidence presented in the primary articles of the MTA.

3. Background

Confusion and controversy might be anticipated from a groundbreaking project as complex as the MTA. One layer of complexity is related to the multiple articles in each of the three sets of publications (i.e., the eight main articles). We will review the findings from each of the three sets of publications in sections here linked to sections in Part II where additional details are provided about the methods of the MTA. For example, in the parallel background section in Part II, we provide additional details about the following methodological topics: (a) genesis of the MTA, (b) design and analysis framework, (c) number, intensity, and duration of treatments, (d) randomization and selection bias, (e) multiple outcome measures, (f) multiple comparisons, (g) adjustment of significance level, (h) multiple definitions of *long term*, and (i) a priori and post hoc tests.

B. Publications at the End of Treatment (MTA Cooperative Group, 1999a, 1999b)

1. Evidence: ITT Analyses

The design of the MTA was based on the principles underlying the RCT approach, which require that participants be randomly assigned to the four treatment conditions: medication management alone (MedMgt), behavior therapy alone (Beh), the multimodal combination of these unimodal treatments (Comb), and treatment "as usual" in the community (CC). Random assignment ensures that the formation of groups is not biased by preferences of those providing or receiving the treatments or by other factors that may operate intentionally or unintentionally. If the sample size is large enough, randomization should produce groups that do not differ in any systematic way at the baseline assessment. The primary analyses of the MTA were based on the ITT approach. This is intended to evaluate the effect of assigned rather than actual treatment. Actual treatment depends on whether the assigned treatments are accepted and implemented as intended. Even if they are not, the primary ITT analyses do not take this into account, although secondary analyses may do so.

The RCT and ITT principles governed the primary analyses of outcomes through the 14-month assessment that produced the initial evidence from the study. We used omnibus tests of treatment by time interactions for multiple outcome measures, followed by multiple pairwise comparisons for 10 outcome measures that showed overall significance. This revealed that multimodal combination (Comb) was not better than medication management alone (MedMgt) on any of these, that MedMgt was better than behavior therapy alone (Beh) on 3 of them, and that community comparison (CC) was worse than Comb on 9, MedMgt on 5, and Beh on none of them.

These analyses reveal two key findings (see MTA Cooperative Group, 1999a). First, assignment to treatment according to the intensive MTA medication algorithm (MedMgt) produced larger benefits (i.e., reduction in ratings of ADHD symptom severity) than assignment to treatment according to the intensive MTA behavior therapy algorithm (Beh). Second, relative

benefit from assignment to the multimodal combination (Comb) was not significantly greater than assignment to the MTA medication algorithm alone (MedMgt).

This evidence based on RCT design and ITT analyses generated some confusion and many controversies, including one based on the interpretation that lack of superiority of the multimodal treatment over medication alone indicated that the intensive behavior therapy provided by the MTA was ineffective. We addressed this erroneous interpretation in secondary analyses.

2. Interpretation: Moderator and Mediator Analyses

Our secondary moderator and mediator analyses were intended to explore whether enhanced response may occur in subgroups and, if so, why this may occur. The findings from these analyses are intended to generate hypotheses for future studies based on speculations on why treatment effects may be enhanced in some cases (the purpose of moderator analyses) or when implemented in particular ways (the purpose of mediator analyses). These exploratory analyses appeal to practical interests of clinicians and encourage speculation about broader questions than those addressed by our narrowly focused primary analyses. Understandably, this may have generated confusion and controversies about the initial findings of the MTA.

We used moderator analyses to search for baseline variables that might identify which participants are likely to have a better response to one treatment than to another. For example, for subgroups defined by socioeconomic status, Comb was statistically superior to MedMgt for the families on public assistance but not for the others with greater family resources, and for subgroups defined by comorbid anxiety, Comb was superior to MedMgt alone (and MedMgt was not significantly better than Beh) for children with comorbid anxiety but not for those without symptoms of anxiety. The moderator analyses have obvious limitations. For example, the use of subgroups increases the number of comparisons that are possible (thus increasing the false-positive probability) and also decreases the sample size for each comparison (thus increasing the false-negative probability). We should point out that statistical significance in one subgroup and not in the other does not prove that the interaction is significant (i.e., that the effect size is statistically different between the two subgroups).We used mediator analyses to search for a change or event that happened during the treatment phase that was correlated with the assigned treatment that may explain all of part of the treatment effect on outcome. This led us to evaluate how treatments are implemented (i.e., acceptance and attendance related to treatment deemed to be “as-intended” or “below-intended”) and to explore how differences that emerged after randomization (i.e., during the course of treatment) might affect outcome. For example, as-intended attendance at the monthly clinic visits during the treatment phase produced better outcome than below-intended attendance, and the adjustments of dose of medication at these visits were different for the Comb and MedMgt groups. By the 14-month assessment point, the Comb group was being treated with a 20% lower dose of medication than the MedMgt group, suggesting that compared to MedMgt, the addition of behavioral treatment in the Comb treatment contributed to the attainment of the same or better results with a lower dose.

3. Qualification

Based on the guidelines of the RCT logic and methodology, our primary ITT analyses were narrowly focused on a few main hypotheses, and the results provide the basic evidence from the MTA (see above). In secondary and many additional articles (see the appendix to Part II), we reported exploratory analyses to supplement the evidence from the primary analyses. In Part II, we provide details about the following topics: (a) additional moderator analyses, (b)

multiple comparisons across outcome domains, (c) single composite outcome measures, (d) evaluation of non-ADHD outcome domains, (e) acceptance of the null hypothesis, (f) commentaries on the end of treatment findings, and (g) the main confusions and controversies at the end of treatment. We will highlight some of these topics here.

We addressed some methodological issues related to the use of multiple outcome measures (i.e., precision of measurement of outcome by any one measure and adjustment of significance level for any one statistical test in order to maintain the nominal significance level across the multiple tests). We explored the use of single outcome measures to increase precision of measurement by compositing multiple measures and to reduce adjustment of the significance level. This was intended to overcome some of the consequences of using multiple outcome measures that may have contributed to lack of statistical significance between treatments. In these articles, we followed the advice of our statistical advisors and focused on estimation of effect sizes rather than tests of statistical significance, but they did show significant superiority of Comb over MedMgt alone on the composite measures (by effect size *d* of .26–.28).

Commentaries about these initial findings of the MTA addressed areas of strength (e.g., large sample, RCT design, successful implementation, and sophisticated analyses) and limitation (the definition of *long term*, the nature of "as usual" treatment, the lack of statistical significance and acceptance of null hypotheses, the use of high-intensity treatments to evaluate efficacy rather than available treatment to address effectiveness, the lack of a placebo control or blind observers, the differences between assigned and actual treatment, and others). We will mention a few important commentaries here that we consider to be controversial and may be based on confusion about the primary findings.

One controversial interpretation of the initial findings was that behavioral treatment was ineffective, which may be due to confusion about relative versus absolute effects in the MTA. In the primary article, the effects of treatment were evaluated by comparison of conditions (relative effects), which were superimposed on a trend of general improvement over time. The first sentence in the discussion section (see MTA Cooperative Group, 1999a) stated that all 4 groups showed marked improvement over time. However, we did not fully explain the obvious overall effect of time, since this could be due to the treatment or to many other unidentified factors (e.g., regression to the mean, expectation effects, secular trends, etc.). This may have contributed to confusion. Instead, we emphasized the interaction of two factors in the regression analysis (treatment and time), and explained why we made this a precondition to be met before performing multiple comparisons of outcomes at the end of treatment phase to contrast treatments and estimate relative benefits. A related controversial interpretation of the initial findings was that multimodal treatment lacked superiority over unimodal treatment with medication, which may be due to confusion about accepting a null hypothesis based on the primary ITT analysis. Our statistical consultants warned us that "the absence of evidence should not be taken as evidence of absence," and we believe that when this wise counsel is not heeded confusion and controversy should be expected, since non-significance cannot be interpreted correctly based on the MTA design and analysis as showing equivalence of the assigned treatments.

C. Publications After the First Follow-Up (MTA Cooperative Group, 2004a, 2004b)

1. Evidence: ITT Analyses

In the primary analyses of outcomes from the 24-month assessment (MTA Cooperative Group, 2004a), we still used the principles of RCT design and ITT analysis. We did this even

though the outcomes evaluated were beyond the end of the treatment phase of the MTA protocol and 10-months after we made individualized recommendations for participants to seek out various clinical services offered in the community. Because the logic and purpose of the RCT approach is to evaluate the long-term consequences of assignment to treatment rather than actual treatment, the use of ITT analysis to produce the primary evidence of the MTA is appropriate even for the followup phase, although it may seem counterintuitive to some. In an observational study such as the MTA follow-up, any observed differences between groups may be due at least in part to factors that influenced which treatments participants selected to get. In the MTA, starting and stopping medication would result in a cross over in terms of actual treatment. So, in the MTA as in any RCT of a chronic condition, when followed naturalistically over time, in the post-treatment phase almost inevitably there is attenuation of effect due to cross-over and heterogeneity of choice within each of the treatment groups.

The narrowly focused ITT analyses revealed a significant effect of the relative impact of assigned treatment on severity of ADHD symptoms ($p < .0001$) at the 24- month assessment. Further comparisons revealed this was because of a persisting relative superiority of the treatments that included assignment to the MTA medication algorithm (Comb and MedMgt) over those that did not (Beh and CC). The magnitude of this comparison of the assigned treatments was reduced by 50% from a moderate effect size (0.6) in the end-of-treatment analyses to a small effect size (0.3) in the follow-up analyses. In the primary article (MTA Cooperative Group, 2004a), current treatment was used as a covariate, and this revealed that across the assigned groups, the participants taking medication at the 24-month assessment point had better outcome than those not taking medication. This was interpreted as an indication that the apparent reduction of the relative superiority of assignment to the MTA medication algorithm (superiority of Comb and MedMgt over Beh and CC) may reflect the lack of maintenance of an effective intervention more than reduction in the effects of medication during the first follow-up phase of the MTA.

The hypothesis of stimulant-related growth suppression had not been evaluated at the end of the treatment phase, so in the first follow-up report (MTA Cooperative Group, 2004b) we provided ITT analyses of measures of physical size at the 14-month assessment point. Thus, in our terminology, this might be considered a primary analysis, despite its delayed and post hoc status. These analyses revealed that the two groups assigned to the MTA medication algorithm had slower growth (i.e., gains of 4.85 cm and 2.53 kg for Comb and 4.25 cm and 1.64 kg for MedMgt) compared to the other two groups (i.e., gains of 6.19 cm and 4.53 kg for Beh and 5.68 cm and 3.13 kg for CC). This represented stimulant-related growth suppression of about 23% in height gain and 47% in weight gain. However, the ITT analysis of growth during the 10-month follow-up revealed the assigned groups did not differ significantly in gain in height or weight (Comb = 5.69 cm and 5.28 kg; MedMgt = 5.69 cm and 5.06 kg; Beh = 6.16 cm and 4.98 kg; CC = 5.79 cm and 4.58 kg). The slight reduced gain in height for the Comb and MedMgt groups provided evidence that there was no growth acceleration (rebound) compared to the Beh and CC groups during the 10-month follow-up.

The extension of the RCT logic and methodology and the use of ITT analyses beyond the treatment phase may have resulted in additional confusion and controversies, including one that suggested the balance of benefits and side effects was different than expected in the literature, with smaller benefits and greater side effects. We addressed this possibility in secondary analyses.

2. Interpretation: Naturalistic Subgroup and Physical Growth Analysis

In the secondary article (MTA Cooperative Group, 2004b), we explored the effects of actual treatment to supplement the analyses of assigned treatment. We explored the use of 14- to 24-month change scores to supplement the primary analyses that evaluated absolute scores at the 24-month assessment point. For these analyses, we formed naturalistic subgroups based on information about actual treatment at the two assessment points (during the 30 days before the 14-month and 24-month assessments). This identified four patterns of treatment over time for those participants who reported consistent use of medication, no use of medication, stopping medication, or starting medication during the follow-up interval. These subgroups were used to evaluate the mediating effect of patterns of treatment over time on efficacy (i.e., ratings of ADHD symptom severity). In the secondary article, an ITT analyses of change scores confirmed that the MTA medication algorithm was associated with deterioration rather than further benefit. Then, a mediator analyses based on the naturalistic subgroups confirmed that the reduction in longterm superiority of the Comb and MedMgt groups relative to the Beh and CC groups may reflect changes in actual use of medication (i.e., lack of maintenance), reflected in the pattern of treatment over time.

One of the main purposes of the secondary article (MTA Cooperative Group, 2004b) was to evaluate whether there were changes in growth velocity during the 14- to 24-month follow-up. As described above, the 24-month ITT analyses revealed that all assigned groups had about the same gain in height during the follow-up interval, suggesting a transitory growth-suppression effect. However, mediator analyses based on our naturalistic subgroups revealed that height and weight gains were less in the always-treated subgroup (4.53 cm and 3.18 kg) than in the never-treated subgroup (5.40 cm and 4.83 kg). A comparison of these patterns suggested that stimulant-related height suppression continued at the same rate (about 1 cm/year reduction in height gain) during the second year when treatment with medication was maintained, but the rate of weight suppression was less (about 1.2 kg/year reduction in weight gain).

3. Qualification

Issues about the impact of changes in the assigned treatments (i.e., starting and stopping medication), which appeared to be relevant for outcome measures related to efficacy (ADHD ratings) and to side effects (height and weight gain), were addressed in both the primary and secondary articles. Multiple exploratory approaches were used to evaluate some obvious hypotheses about persistence and desistance of both efficacy (e.g., reduction of ADHD symptom severity) and side effects (stimulant- related growth suppression). In Part II, we provide additional details about two topics: (a) partial loss of relative benefits of medication and (b) the main confusions and controversies at the first follow-up.

Even in the primary article (MTA Cooperative Group, 2004a), mediator analyses were performed to evaluate how changes in medication use between the 14- and 24- month assessment points might have affected outcomes. At the 14-month assessment, compliance with assigned medication use was high (87% and 93%) in the Comb and MedMgt groups, but this decreased by the 24-month assessment (to 70% and 72%). In contrast, in the Beh and CC groups, the percentage of participants with medication use increased from 23% and 55% at the 14-month assessment to 38% and 62% by the 24-month assessment. By adding a covariate (interim medication use) in the ITT analyses, we documented that the effect of current medication use was statistically significant. This suggested that the reduction in relative superiority of the MTA medication algorithm for symptomatic improvement was related to lack of maintenance of medication. However, even when adjusted for the beneficial effects of current medication use, the effect of assigned treatment was still significant. This indicated that the persisting effect of

the relative superiority of assignment to the Comb and MedMgt conditions versus the Beh or CC conditions was not entirely dependent on continuing the use of medication (i.e., that there might be a positive "carry-over" effect even when medication use was stopped).

This pattern was confirmed in the secondary article, which evaluated 14- to 24-month change scores. The participants who stopped medication deteriorated, whereas those who started medication improved during the 10- month follow-up. Based on this, we speculated that the patterns of actual treatment that emerge over time may explain most of the observed partial loss of the relative benefits of initial assignment to the MTA medication algorithm. We also predicted that if this trend of actual use of medication continued, then eventually there would be a complete loss of the effect of assigned treatment evaluated by ITT analysis.

The analyses of side effects at the 24-month assessment point suggested that the growth suppression effect was attenuated, and the contrast of assigned groups was no longer statistically significant. This suggests that differential treatment effects on growth dissipated during followup. However, there was no evidence of growth rebound in the Comb and MedMgt groups relative to the Beh and CC groups during the 10-month follow-up (i.e., there was no "catch-up"). Also, a mediator analysis was performed based on subgroups defined by patterns of actual treatment over time. This exploratory analysis suggested that the participants who were continuously medicated showed slower growth compared to those who were not medicated, suggesting that stimulant-related growth suppression may continue when treatment is maintained.

Thus, at this point in the MTA follow-up (at the 24- month assessment), both the primary and secondary articles suggested that the outcomes over time depended on maintenance of treatment with medication and actual treatment patterns. This held for measures of long-term efficacy (reduction of severity of symptoms) as well as a measure of long-term side effects (reduction in height gain).

D. Second Follow-Up (Jensen et al., 2007; Swanson, Elliott, et al., 2007; Swanson, Hinshaw, et al., 2007; Molina et al., 2007)

1. Evidence: ITT Analyses

The continued use of the RCT logic and methodology again led us to perform narrowly focused ITT analyses to evaluate the effects of assigned rather than actual treatment through the 36-month assessment (Jensen et al., 2007). This primary analysis revealed a further decrease in the initial large relative advantage of assigned treatment with stimulant medication on ratings of ADHD symptom severity, and by this point in the MTA follow-up, the relative superiority (e.g., greater reduction in ADHD symptom severity related to assignment to Comb and MedMgt compared to Beh and CC) was completely lost. The ITT analyses indicated that the assigned groups did not differ significantly on any of five outcome measures, and a comparison of effect size for the most sensitive measures for treatment effects (i.e., a composite of parent and teacher ratings of severity just for ADHD symptoms) indicated that the large relative superiority of the MTA medication algorithm at the 14-month assessment (.86) was negligible at the 36-month assessment (.10).

We were concerned that the continued use of the RCT approach and ITT analyses might create new confusions and controversies, so mediator analyses were reported in the primary article (Jensen et al., 2007) to evaluate whether the loss of relative superiority of medication was relative to maintenance of treatment, as had been predicted. However, we could not show that the loss was due to a further decrease in the percentage of Comb and MedMgt cases using

medication, which remained about the same at the 24- and 36-month assessments. Instead, analysis of 24- to 36-month change in treatment and outcome indicated that continued medication use was a marker of deterioration rather than benefit. We addressed this observation in secondary analyses.

2. Interpretation: Multiple Secondary Articles

After the first follow-up, we hypothesized that a continued loss of relative superiority of the Comb and MedMgt groups would be associated with continued decrease in the actual use of medication in these groups, but this was not confirmed in the analyses of the second follow-up. Instead, the continued loss occurred despite maintenance of about the same level of medication use during the second follow-up phase of the MTA. Thus, the long-term follow-up over the 3-year period suggested that the relative benefits of early use of the intensive medication management approach of the MTA might be temporary (i.e., that the initial relative superiority may not last forever). Three secondary articles accompanied the primary article to address this finding, and we will discuss each of these here (see below).

Swanson, Hinshaw, et al. (2007) addressed the hypothesis that the loss of relative superiority may be due to selection bias (i.e., the selective treatment of the most severe cases, which might operate to mask underlying benefits of medication). This was evaluated using the propensity score method. Despite the use of this sophisticated statistical method, our findings did not support this key hypothesis. Another approach addressed whether individual differences in response may have masked beneficial effects of medication. This was evaluated using the statistical method of growth mixture model analysis. This complex analysis identified three latent classes (subgroups) with different trajectories of outcome (e.g., decreases in ADHD symptom severity) over time. In two of the classes, there was a large initial improvement (decrease in ratings of ADHD symptoms severity). In one of these (52% of the sample), the large initial large improvement was maintained over time, but in the other (14% of the sample), it dissipated over time. The trajectory of another class (34% of the sample) was defined by initially modest but gradually greater improvement over time. The relative advantage of medication was small but significant at the 14-month assessment and then increased over time and remained significant at the 36-month assessment. Thus, in this subgroup, those participants taking medication fared significantly better than those not taking medication, and the relative benefit did not dissipate over time. However, in the other two classes (comprising 66% of the total sample), a different pattern was observed. A large and significant medication effect was present initially but then did dissipate over time. Apparently, this effect (dissipation of the beneficial effects of medication) that characterized most of the participants (66%) overshadowed the longterm beneficial effect of medication observed in a smaller subset of participants (34%).

Swanson, Elliott, et al. (2007) extended the evaluation of patterns of actual medication in naturalistic subgroups defined by patterns of treatment over time. Information about treatment prior to entry into the MTA as well as during the MTA was used, which provided four time points for medication status (at the baseline, 14-month, 24-month, and 36-month assessments). This allowed us to establish subgroups for participants who were never treated, always treated (even before entering the MTA), newly treated (continuously after entering drug naïve), and inconsistently treated (because of starting or stopping) with medication. We hypothesized that consistent treatment with medication might result in maintenance of superiority at the 36-month assessment. However, this was not confirmed: Our naturalistic subgroups did not differ significantly on ADHD symptom-severity at the 36-month assessment. In addition, we hypothesized that growth rebound might occur, but this was not observed either. Those whose

height trajectory had been slowed by medication use did not show further reduction in height gain, but they did not "catch up" (i.e., they did not make up the prior less-than-expected height gain with a subsequent greater-than-expected height gain).

Molina et al. (2007) evaluated new outcome domains that become relevant during early adolescence, including substance use experimentation and juvenile delinquency. We used ITT analyses to test the widely held hypothesis that early treatment with stimulant medication (i.e., assignment to the MTA medication algorithm) would provide protection from these adverse outcomes during this critical period of development from childhood to adolescence. However, the results did not support this hypothesis. Instead, the only statistically significant effect suggested protection against early substance use in groups assigned to treatments including the MTA behavior therapy algorithm (Comb and Beh) compared to those that did not (MedMgt and CC).

The findings of these secondary articles were not consistent with some expectations about medication effects that were generally accepted by many of the field's investigators and clinicians in 2007. That is, long-term benefits from consistent treatment were not documented; selection bias was not shown to account for the loss of relative superiority of medication over time; there was no evidence for "catch-up" growth; early treatment with medication did not protect against later adverse outcomes. We expect that these challenges to the field's views will contribute to future controversies about the long-term outcomes in the MTA. For example, after the 14-month treatment phase, all MTA participants received treatment "as usual" in the community. We and others have speculated that changes in how treatment with medication was managed may have contributed to loss of relative benefits over time. We were not able to document this, but it remains as a possibility that should be evaluated in future research.

3. Qualification: Challenges to Consensus Views and New Predictions

The set of four articles describing the second followup were published so recently that there has not been sufficient time yet for subsequent articles or commentaries to be published. Therefore, in this section, we provide some commentary to initiate discussion about what they suggest—that is, the lack of long-term effects of past or current treatment with stimulant medication. In Part II, we provide additional details about two topics: (a) the main confusions and controversies at the second followup and (b) subsequent studies of multimodal treatment of children with ADHD.

First, it is important to express caution about interpretations based on observations of nonsignificance of statistical tests. Based on the limitations of the design and methods of the MTA, the dangers of acceptance of the null hypothesis should be noted, and additional alternative hypotheses should be considered. Second, it is important to remember that the evaluation of actual treatment as well as assigned treatment is an important part of the follow-up analyses. With these and other cautions in mind, we will point out some qualifications of the three secondary articles about outcomes at the 36-month assessment.

The hypothesis of selection bias cannot be dismissed (Swanson, Hinshaw, et al. 2007) by propensity score analysis because the propensity score is based on the finite set of measures we established with our MTA battery. Therefore, many other unmeasured factors could be operating to mask underlying benefits of treatment with stimulant medication. Also, the use of propensity score analyses rests on the assumption that selection biases can be modeled as a simple linear combination of multiple variables (e.g., severity of symptoms, previous experience with medication, initial treatment response, ethnicity, socioeconomic status, and other variables), and this assumption may not hold for the complex baseline measures of the MTA battery.

The hypothesis of growth rebound cannot be dismissed yet (Swanson, Elliott, et al., 2007), because it is still possible that rebound may occur during adolescence. This is most relevant for height, because weight and body composition are subject to variation at any stage of development but height is not. Logically, the hypothesis of growth rebound for height would not be adequately evaluated until ultimate height is attained in adulthood. The evaluations at the 36-month assessment were conducted when most of the MTA participants were between 11 and 13 years of age. The outcomes at the 12-year assessment (when the participants will be 19 to 21 years of age) will provide more definitive data to address this important issue. Also, some uncertainty about the effects of prior treatment exists, because the largest effect was due to a baseline difference between the always- and never-treated subgroups, even though these subgroups were established prospectively by patterns of medication use during the MTA.

Our formation of homogeneous subgroups with growth mixture model analysis was based on the limited trajectory of outcome with just four assessment points from baseline through the 36-month assessment (Swanson, Hinshaw, et al., 2007). The heterogeneity of trajectories in the sample could change during adolescence and into adulthood, so the statistically defined subgroups may change in future evaluations of long-term outcome. The reported patterns observed so far, with two subgroups (66% of the sample) showing possible decreasing effectiveness of medication and one (34% of the sample) showing possible increasing effectiveness, may change when outcomes at additional assessment points are included in the growth mixture model analyses. Subsequent analyses based on this technique will be able to determine whether these patterns continue, or whether different patterns across the same or different subgroups emerge when the sample is older and at a more mature stage of development.

The hypothesis of the protective (or predisposing) effects of early treatment cannot be dismissed (Molina et al., 2007) at this time, because it cannot be adequately or fully tested until adolescent and adult patterns of substance abuse and dependence emerge to be evaluated. The growth mixture model analysis of these outcome measures did not identify heterogeneity of trajectories over time, so subgroups were not defined and evaluated as in the growth mixture model analysis of the ADHD symptom-severity outcome measure (see above). Based on a developmental perspective, this was not unexpected because the range of outcomes on these variables is low until adolescence, and some measures of substance use and antisocial behavior are not expected to become stable until adulthood. Future analyses will address these developmental issues.

D. Clinical Relevance: Practical Suggestions and Future Directions

The initial findings of the MTA at the 14-month assessment (MTA Cooperative Group, 1999a, 1999b) clearly show relative superiority of assignment to the MTA medication algorithm on ADHD symptoms. This provides evidence of long-term benefits of stimulant medication for over 1 year. This provided an answer to the fundamental question the MTA was designed to address. These initial findings were important, because when the MTA was initiated there was little information in the literature on the long-term effects of stimulants.

The findings from the first follow-up provided another contribution to the sparse literature on long-term effects of medication. The statistically significant persistence at the 24-month assessment provides evidence of long-term benefits over 2 years. The observation of partial loss of the superiority of assignment to the MTA medication algorithm should be interpreted in light of maintenance of treatment, which was reduced in the transition to treatment “as usual” in the community and some documented limitations of community medication treatment.

The findings from the second follow-up at the 36- months assessment suggest the temporary nature of the superiority of assignment to intensive, carefully monitored medication management (i.e., to the MTA medication algorithm), which may gradually dissipate completely when the children are returned to community treatment. These interim findings also suggest that the relative benefits are not apparent even in subgroups of participants who maintained treatment over time (i.e., when comparing the subgroup of participants never treated to the subgroup always treated with stimulant medication).

The findings at successive follow-up phases of the MTA do not contradict the prior findings from an earlier phase, but at each step, so far, the findings have provided new perspectives to be considered about long-term effects of treatment of children with ADHD.

In Part II, we provide additional details about the following questions: (a) What is the first line treatment for ADHD? (b) Do stimulants suppress physical growth? (c) Does recommendation of medication provide long-term benefit? (d) Does maintenance of treatment produce maintenance of relative superiority of medication? (e) Does selection bias mask beneficial effects of medication? and (f) Do some subgroups benefit more than others?

Acknowledgements

The Multimodal Treatment Study of Children with ADHD (MTA) was funded by the National Institute of Mental Health (NIMH) as a cooperative agreement involving six clinical sites. Collaborators from the National Institute of Mental Health are Peter S. Jensen, MD (currently at the REACH Institute), L. Eugene Arnold, MD, MEd (currently at Ohio State University), Joanne B. Severe, MS (Clinical Trials Operations and Biostatistics Unit, Division of Services and Intervention Research), Benedetto Vitiello, MD (Child & Adolescent Treatment and Preventive Interventions Research Branch), Kimberly Hoagwood, PhD (currently at Columbia University); previous contributors from NIMH to the early phase: John Richters, PhD (currently at National Institute of Nursing Research); Donald Vereen, MD (currently at National Institute on Drug Abuse). Principal investigators and co-investigators from the clinical sites are from University of California, Berkeley/San Francisco: Stephen P. Hinshaw, PhD (Berkeley) and Glen R. Elliott, PhD, MD (San Francisco); Duke University: C. Keith Conners, PhD, Karen C. Wells, PhD, John March,MD, MPH, and Jeffery Epstein, PhD; University of California, Irvine/Los Angeles: James Swanson, PhD (Irvine), Dennis P. Cantwell, MD, (deceased, Los Angeles), and Timothy Wigal, PhD (Irvine); Long Island Jewish Medical Center/Montreal Children's Hospital: Howard B. Abikoff, PhD (currently at New York University School of Medicine) and Lily Hechtman, MD (McGill University); New York State Psychiatric Institute/Columbia University/Mount Sinai Medical Center: Laurence L. Greenhill, MD (Columbia University) and Jeffrey H. Newcorn, MD (Mount Sinai School of Medicine); University of Pittsburgh: William E. Pelham, PhD (currently at State University of New York, Buffalo), Betsy Hoza, PhD (currently at University of Vermont), and Brooke Molina, PhD. The original statistical and trial design consultant was Helena C. Kraemer, PhD (Stanford University), and the follow-up phase statistical collaborators are Robert D. Gibbons, PhD (University of Illinois, Chicago), Sue Marcus, PhD (Mt. Sinai College of Medicine), and Kwan Hur, PhD (University of Illinois, Chicago). Additional collaborators are Thomas Hanley, EdD from the Office of Special Education Programs/US Department of Education and Karen Stern, PhD from the Office of Juvenile Justice and Delinquency Prevention/Department of Justice.

Address correspondence to James Swanson, UCI Child Development Center, 19722 MacArthur Blvd., Irvine, CA 92612; jmswanso@uci.edu. The work reported was supported by cooperative agreement grants and contracts from the National Institute of Mental Health to the following: University of California, Berkeley: U01 MH50461 and N01MH12009; Duke University: U01 MH50477 and N01MH12012; University of California, Irvine: U01 MH50440 and N01MH 12011; Research Foundation for Mental Hygiene (New York State Psychiatric Institute/Columbia University): U01 MH50467and N01 MH12007; Long Island-Jewish Medical Center U01 MH50453; New York University: N01MH 12004; University of Pittsburgh: U01 MH50467 and N01 MH 12010; McGill University N01MH12008. The Office of Special Education Programs of the U.S. Department of Education, the Office of Juvenile Justice and Delinquency Prevention of the Justice Department, and the National Institute on Drug Abuse also participated in funding.

The opinions and assertions contained in this report are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of Health and Human Services, the National Institutes of Health, or the National Institute of Mental Health.

References

Jensen, P. S., Arnold, L. E., Swanson, J. M., Vitiello, B., Abikoff, H. B., Greenhill, L. L., et al. (2007). 3-year follow-up of the NIMH MTA study. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*, 989-1002.

MTA Cooperative Group. (1999a). A 14-month randomized clinical trial of treatment strategies for attention deficit hyperactivity disorder. *Archives of General Psychiatry, 56*, 1073-1086.

MTA Cooperative Group. (1999b). Moderators and mediators of treatment response for children with attention-deficit/hyperactivity disorder. *Archives of General Psychiatry, 56*, 1088-1096.

MTA Cooperative Group. (2004a). National Institute of Mental Health Multimodal Treatment Study of ADHD follow-up: 24-month outcomes of treatment strategies for attention-deficit/hyperactivity disorder (ADHD). *Pediatrics, 113*, 754-761.

MTA Cooperative Group. (2004b). National Institute of Mental Health Multimodal Treatment Study of ADHD follow-up: Changes in effectiveness and growth after the end of treatment. *Pediatrics, 113*, 762-769.

Molina, B. S., Flory, K., Hinshaw, S. P., Greiner, A. R., Arnold, L. E., Swanson, J. M., et al. (2007). Delinquent behavior and emerging substance use in the MTA at 36 months: Prevalence, course, and treatment effects. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*, 1028-1040.

Swanson J. M., Elliott, G. R., Greenhill, L. L., Wigal, T., Arnold, L. E., Vitiello, B., et al. (2007). Effects of stimulant medication on growth rates across 3 years in the MTA follow-up. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*, 1015-1027.

Swanson, J. M., Hinshaw, S. P., Arnold, L. E., Gibbons, R. D., Marcus S., Hur, K., et al. (2007). Secondary evaluations of MTA 36-month outcomes: Propensity score and growth mixture model analyses. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*, 1003-1014.